

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE **8/27/04**

3. REPORT TYPE AND DATES COVERED
Final 06/01/99-05/31/04

4. TITLE AND SUBTITLE

Sensors & Symbols: An Integrated Framework

5. FUNDING NUMBERS

N00014-99-1-0464

6. AUTHOR(S) **Professor Daphne Koller**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

**Stanford University School of Engineering
Computer Science Department
GATES BLDG 1A-142
Stanford, CA 94305-9010**

8. PERFORMING ORGANIZATION
REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

**Wendy L. Martinez
Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington, VA 22217-5660**

10. SPONSORING / MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The goal of this effort was to provide a unified probabilistic framework that integrates symbolic and sensory reasoning. Such a framework would allow sensor data to be analyzed in terms of high-level symbolic models. It will also allow the results of high-level analysis to guide the low-level sensor interpretation task and to help in resolving ambiguities in the sensor data. Our approach was based on the framework of probabilistic graphical models, which allows us to build systems that learn and reason with complex models, encompassing both low-level continuous sensor data and high-level symbolic concepts. Over the five years of the project, we explored two main thrusts:
Inference and learning in hybrid and temporal Bayesian networks
Mapping and modeling of 3D physical environments

Our progress on each of these two directions is detailed in the attached report.

20041008 463

14. SUBJECT TERMS

15. NUMBER OF PAGES

10

16. PRICE CODE

ONR Grant N00014-99-1-0464 — Final Report

PI: Professor Daphne Koller, Stanford University

The goal of this effort was to provide a unified probabilistic framework that integrates symbolic and sensory reasoning. Such a framework would allow sensor data to be analyzed in terms of high-level symbolic models. It will also allow the results of high-level analysis to guide the low-level sensor interpretation task and to help in resolving ambiguities in the sensor data. Our approach was based on the framework of probabilistic graphical models, which allows us to build systems that learn and reason with complex models, encompassing both low-level continuous sensor data and high-level symbolic concepts.

The award spanned five years of work, and therefore covered a range of activities related to this overall goal. Below we summarize our accomplishments for each of these activities, proceeding roughly in chronological order.

1 Hybrid and Temporal Bayesian Networks

In one thrust, we developed fundamental methods for reasoning in probabilistic graphical models. Our focus was on extending the capabilities of these models in two directions that we viewed as critical to the project direction: models that represent processes that evolve over time, and models that involve both discrete and continuous variables.

1.1 Temporal Models

When tracking a complex high-dimensional system, we must keep track of a distribution over an exponentially large state space. In [7, 5], we showed that, in compound systems composed of interacting subsystems, the correlations between the state of different subsystems can be quite weak. Thus, we can often provide a high-quality approximation to a complex belief state by approximating the distribution using simpler factors, e.g., independent marginals over subsystem states. The main advantage of this approach is the ability to keep many distinct hypotheses in a compact representation.

In a parallel effort [6, 4], we considered the problem of efficiently learning temporal probabilistic models from partially observable data. This task was important for the project, as many variables (particularly symbolic ones such as intention) are rarely observed. It is also a very difficult task from a computational perspective. We have investigated both the problem of parameter learning and of learning model structure — the dependence of one variable on others — from partially observable data. Building on some of our previous work, we have substantially improved the performance of algorithms for these tasks. We have also shown how these techniques can be applied online, as the model is being used to track the world. We also showed promising results on the very challenging problem of inducing the existence of hidden variables automatically from the data.

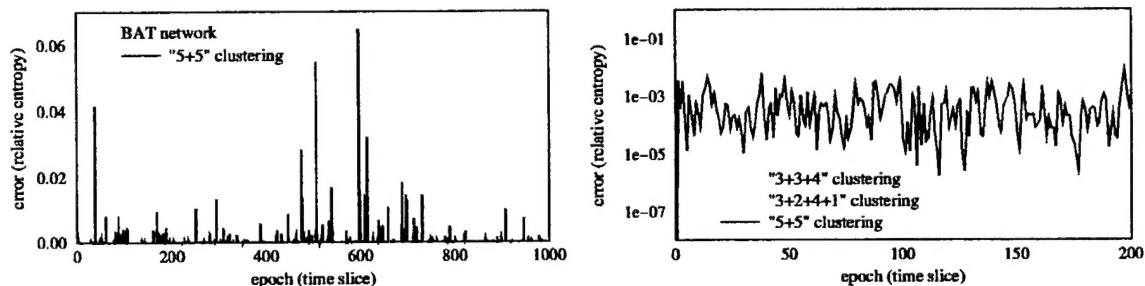


Figure 1: Sample experimental results for approximate tracking algorithm: (a) relative entropy error of approximate tracking algorithm for a typical run using a freeway driving network; (b) comparison of relative entropy error for three different approximate belief state representations; the speedup in running time is approximately a factor of 30.

1.2 Hybrid Models

Up to a few years ago, most work on probabilistic graphical models has focused on representing distributions involving only discrete or only continuous variables. A large component of our effort focused on extending this framework to allow representation and reasoning with a rich class of models involving both discrete and continuous variables, and dependencies between them. Our work on this topic resulted in several contributions.

1.2.1 Anytime Inference in CLG Networks

We first studied the important case of conditional linear Gaussian (CLG) networks, where the continuous nodes all have linear Gaussian models, and discrete nodes cannot have continuous parents. Under these conditions, the conditional distribution of the continuous variables given the discrete ones is a multivariate Gaussian, and the distribution as a whole is an exponentially large mixture of Gaussians.

There are known algorithms for inference in such networks, which are very similar in spirit to inference algorithms for discrete networks. Thus, there was a common perception that inference in CLG networks was “about the same” as inference in discrete networks. In [9], we proved that this perception is false: Even for network structures for which inference in the discrete case is very easy (linear time), inference for CLGs can still be NP-hard. In particular, we showed that even in an extremely simple class of CLGs, where the network structure is a polytree and every continuous variable has at most one (binary) discrete ancestor, inference is NP-hard. An even more surprising result shows that, unless $P=NP$, there does not exist a polynomial time approximate inference algorithm with absolute error smaller than 0.5. It is important to note that these results also apply to the popular model of Switching Kalman Filters; thus, we provided the first formal complexity results for this important class of models.

Given that inference for very simple CLGs is NP-hard, and that even approximate inference is not tractable, one might conclude that inference in CLG models is a lost cause. Fortunately, many real life domains have special features which can be exploited by efficient algorithms. Specifically, it is often the case that a relatively small subset of Gaussians is a good approximation for the entire mixture. In models of physical systems, scenarios involving multiple simultaneous faults are very unlikely. In physical tracking domains, hypothesis involving very frequent changes in destination or activity are rare. Our key idea is to consider the unlikely hypotheses only if the likely hypotheses

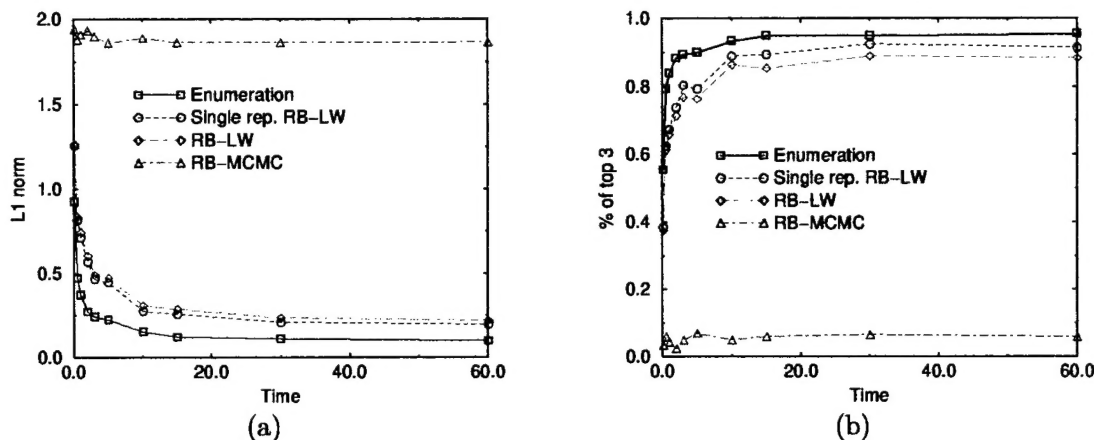


Figure 2: Results of various hypothesis enumeration algorithms on random networks (a) L1 norm to correct distribution (b) Percentage of runs where the most likely hypothesis generated by the algorithm is one of the three most likely assignments.

fail to explain the evidence. Our algorithm works by enumerating the assignments to the discrete variables in decreasing order of prior likelihood.¹ The cost of this algorithm grows with the cost of inference on the discrete part of the model, which is, in almost all cases, significantly easier than inference over the hybrid model.

1.2.2 Nonlinear models

In a second direction, we relaxed the very strong restriction to CLG networks, which assumes linear dynamics, and disallows dependencies of discrete variables on continuous ones.

In one direction [8], we considered the case of nonlinear continuous dynamics. Unlike many standard approaches, we did not linearize the dynamics. Rather, given a nonlinear dependence of a variable X on its parents Y_1, \dots, Y_k , we generate the joint distribution over X, Y_1, \dots, Y_k , and then use numerical integration techniques to find a good Gaussian approximation to this joint distribution by computing the relevant expectations. The dimensionality of the integrals involved is k . In practice, the structure of the BN representation results in fairly localized dependencies, so that k is quite small. Our results [8] show that this method is highly effective, and significantly more accurate (per time spent) than other approaches.

A second type of nonlinear dynamics is induced by the dependencies of discrete variables on continuous ones. Unfortunately, there is no exact inference algorithm known for such networks. One can always resort to the use of approximate inference, such as discretization or sampling, but these approaches have some serious limitations. It is often hard to find a good discretization: Sometimes any reasonable discretization demands too fine a resolution, and often requires the handling of intractable intermediate factors (especially in high dimensions). The convergence of sampling algorithms can be quite slow, and is very sensitive to the network parameters and the configuration of the evidence.

¹Note that, when applying this algorithm to a dynamic problem, this prior likelihood will depend on evidence obtained in previous time steps, and hence will be a fairly good indicator.

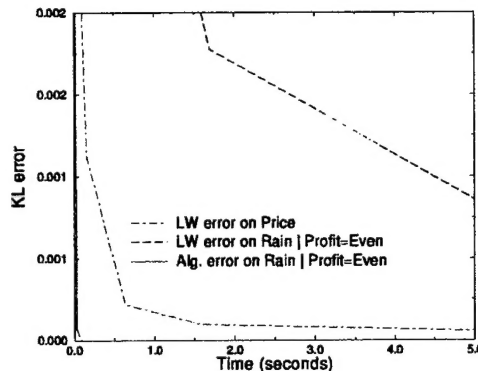


Figure 3: Comparison of our approach to non-linear CPDs with Likelihood Weighting. The error of our algorithm on the *Price* query is 0.

In [11], we developed the first “exact” algorithm for the class of *augmented CLG networks*, which are networks where the continuous variables have a CLG model (as described above), but that allow a logistic (softmax) dependency of the discrete variables on the continuous ones. Our algorithm is based on the simple idea of approximating the product of a Gaussian and a softmax as a Gaussian, where we construct the approximation using numerical integration. We show how we can embed this approach within the general framework of Lauritzen’s algorithm for inference in standard CLG networks. The resulting algorithm is quite simple, and often comparable in its complexity to Lauritzen’s original algorithm. Our algorithm is exact in a sense that is analogous to Lauritzen’s algorithm: It computes the exact distributions over the discrete nodes, and the exact first and second moments of the continuous ones, *up to inaccuracies resulting from numerical integration used within the algorithm*.

1.2.3 Dealing with the temporal blowup

The techniques described above allow us to do inference in a static network. We then proceeded to use this algorithm as a key subroutine in a dynamic tracking algorithm. Given a mixture distribution that approximates our belief state at time t , we can use it to generate a new mixture distribution that approximates our belief state at time $t+1$. However, we cannot afford to propagate every mixture component into time $t+2$, as that would exponentially increase the size of the mixture over time, and therefore we need to reduce the size of the mixture. The problem is that it is very difficult to determine which hypotheses to keep for complex systems. Naively, we would choose to keep the most likely hypotheses and remove the others. However, sometimes a crucial piece of information is not manifested until several time steps after the hypotheses are pruned. The danger is that, without supporting evidence, the correct hypothesis would be removed from our belief state.

Our approach to dealing with this problem, as described in [10], has three components. We first use an algorithm that collapses similar hypotheses into a single hypothesis. This algorithm is combined with a novel approximate smoothing algorithm that we use to improve our ability to find the more likely hypotheses. Finally, we combine our techniques with a decomposition method based on our earlier work for discrete networks [7] that allows the tracking of very large systems that involve many possible failures in different components.

Our collapsing algorithm is based on the observation that, among the likely hypotheses, we often have very similar ones. For example, hypotheses which correspond to measurement faults 20 steps ago and 21 steps ago would often be almost identical. Instead of keeping similar hypotheses in our belief state we can collapse them into one and use the remaining slots to keep other, distinctly different, hypotheses. The main question is the choice of which hypotheses to collapse. We provide a novel approach that takes into consideration both the likelihood of the different hypotheses and their similarity to each other.

Hypothesis collapsing is a myopic method; it only uses evidence observed up to time t . As discussed above, in some cases the likelihood of the current hypothesis only increases after a certain delay. As our approach is more likely to collapse hypotheses that are currently unlikely, the correct hypothesis might be lost before it has a chance to reveal itself. The obvious solution to the problem is to pick the likely hypotheses based not only on past and present evidence but also on future evidence. However, to do so, we must first propagate a belief state forward in time, and this is the very problem we are trying to solve. We break this cycle by using a simpler method of collapsing hypotheses, and then performing a backward propagation process only for the hypothesis weights. We use the more informed hypothesis weights as the basis for our collapsing algorithm.

Finally, none of these approaches are sufficient to deal with very complex high-dimensional systems. In this case, the number of hypotheses required to appropriately represent the belief state can grow extremely large. We use a continuous extension to our work in [7, 5], to represent the belief state as a combination of simpler factors over subsystem states.

1.3 Application to Diagnosis

We tested our methods on a complex real-world task involving both discrete and continuous variables. This task was the diagnosis of a real system constructed by NASA. The system, called *RWGS* (*Reverse Water Gas Shift*), converts carbon dioxide and hydrogen into oxygen, for the purpose of eventually making fuel on Mars. The diagnosis task involves both discrete variables, corresponding to different types of component failures, and continuous ones, representing the (hidden) system state and the sensor measurements.

The RWGS presents a number of significant modeling and algorithmic challenges. From a modeling perspective, the system is very complex, and contains many subtle phenomena that are difficult to model accurately. Various phenomena in the system manifest themselves over dramatically different time scales, ranging from pressure waves that propagate in a time scale of milliseconds to slow changes such as gas composition that take hours to evolve. From a tracking perspective, the system dynamics are complex and highly nonlinear. Furthermore, the sensors give only a limited view of the system state. Some key quantities of the system are not measured, and the available sensors are noisy and biased, with both the noise level and the bias varying with the system state.

We constructed a probabilistic graphical model for this problem, where the state at each time point is represented by 8 discrete and 176 continuous variables. We showed [8] how the methods developed in our work allow us to deal effectively with the challenges involved in such a system. In addition to the contributions described above, we also showed how to use a fixed-point computation to deal with effects that develop at different time scales, specifically rapid changes occurring during slowly changing processes. Our results showed the ability of our methods both to track the continuous state of the system, and to provide accurate conclusions regarding the hidden discrete state of the system (the underlying failure modes).

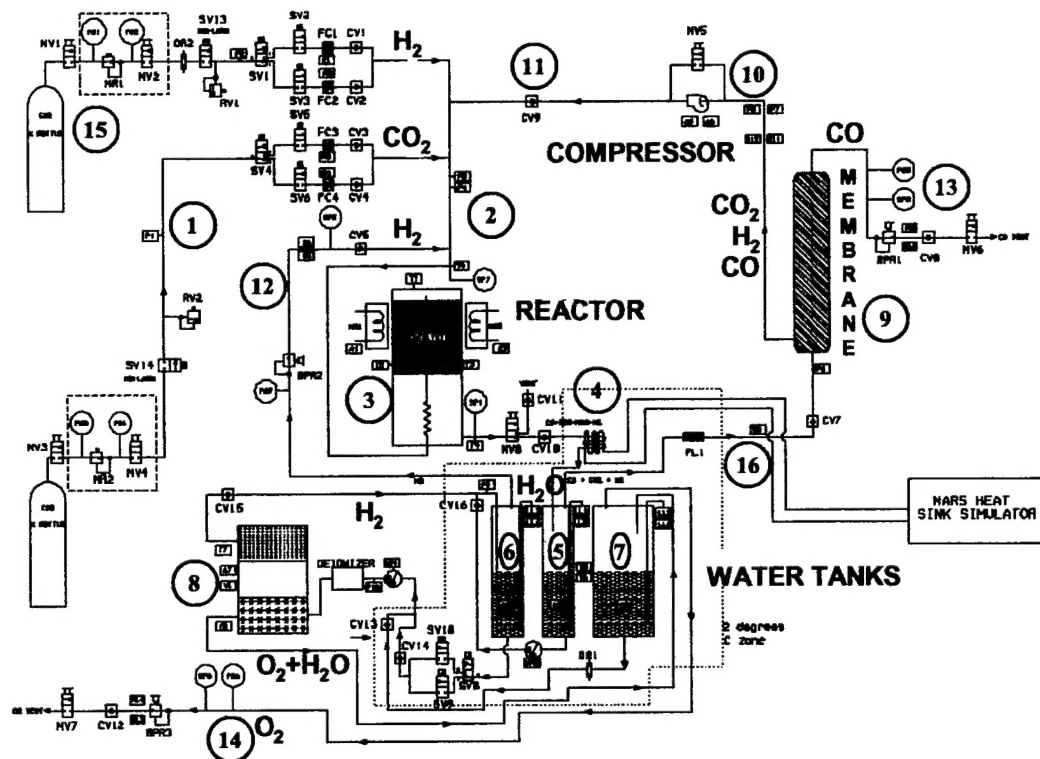


Figure 4: The RWGS schematic

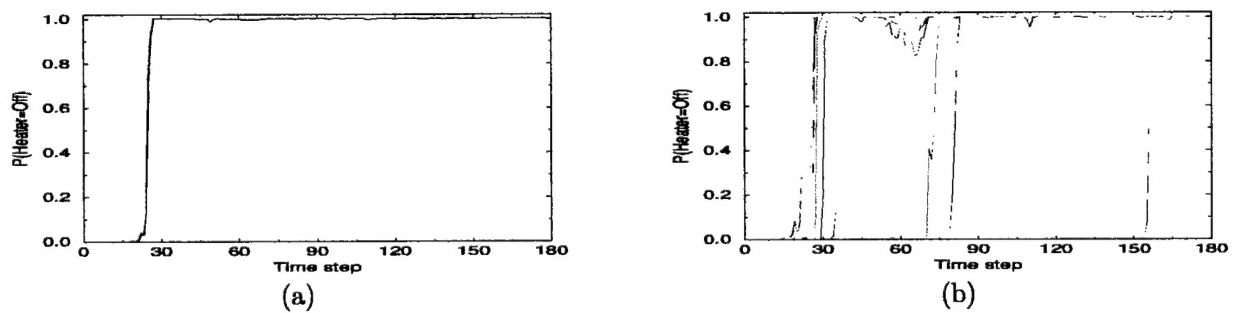


Figure 5: Heater shutdown scenario: likelihood of heater shutdown for (a) our algorithm, (b) Different runs of the sampling-based Rao-Blackwellized Particle Filtering algorithm. The actual shutdown occurs at time step 11.

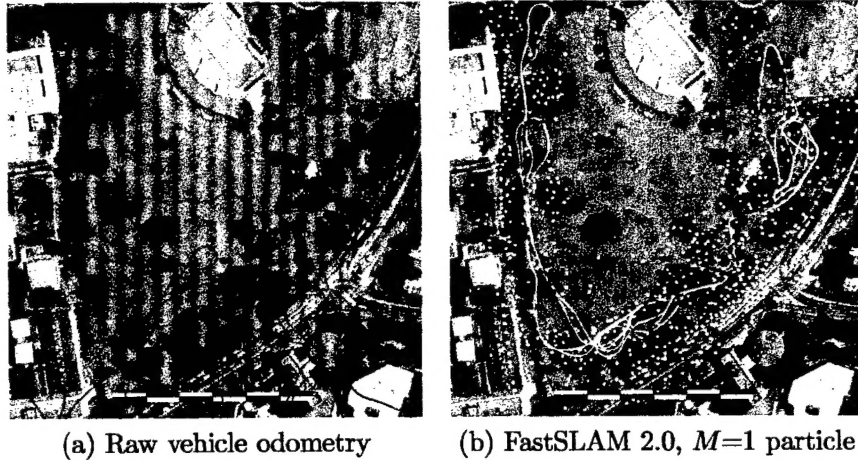


Figure 6: FastSLAM 2.0 algorithm applied to the Victoria Park benchmark data set using only a single particle. The accuracy of the recovered path and the resulting map is indistinguishable from that the best EKF-style methods and the original FastSLAM algorithm with $M=100$ particles.

2 Symbolic Maps of Physical Environments

In the second phase of the project, we focused on the goal of constructing a symbolic map of an unknown environment from raw robot sensor data. For example, consider a robot scanning an office environment. It will observe several objects such as desks, monitors, chairs and people. An ultimate goal would be to recognize and segment out the objects in the office, creating a map containing symbolic descriptions such as “this is a chair next to the door.” To do so, we need to deal with some key problems, involving basic mapping in unknown environments, constructing maps at the object level, and building shape models of physical objects. In collaboration with Professor Sebastian Thrun (who has recently moved from Carnegie Mellon University to Stanford), we have explored these three important problems.

2.1 SLAM

Our first focus was on the SLAM task — Simultaneous Localization And Mapping — mapping an unknown environment using a robot whose trajectory through the environment is also unknown. This problem is a key problem in robotics, and is a critical component in any system that tackles the mapping problem in a real-world setting. We explored several different approaches to this problem, all based on our main research theme of exploiting the structured representation of probabilistic graphical models.

One of our proposed algorithms [12, 13] exploits important independence structure in the SLAM problem: the fact that the landmark positions are conditionally independent given the robot’s motion path. We use particle filtering methods to sample the robot’s path, and then a set of low-dimensional extended Kalman filters to represent our beliefs regarding the position of each landmark given each robot’s path. This method allows the SLAM problem to be solved for significantly larger and more complex environments than have been addressed so far. Moreover, it can be extended to provide a new solution to the data association problem, where the robot is uncertain about which landmark in the environment it is sensing. In real environments, it is rarely the case that landmarks are always uniquely identifiable, so that algorithms that address this problem are of

practical importance. We prove convergence of this new algorithm for linear SLAM problems and provide real-world experimental results that illustrate an order of magnitude improvement over other solutions to SLAM.

In a different algorithm focused on solving the linear SLAM problem in very high-dimensional spaces [14], we employ techniques that are very similar to our early work on tracking in discrete systems [7]. This method represents our belief state — the distribution over the current map and robot pose — using a probabilistic graphical model that shows the correlations in our beliefs about the positions of different landmarks. Over time, the network becomes densely connected, so that we approximate the belief state by ignoring weak correlations between the landmarks. We show that the different update steps in this algorithm can be executed (approximately) in constant time, irrespective of the size of the map. We also provide empirical results obtained for a benchmark data set collected in an outdoor environment, and using a multi-robot mapping simulation.

2.2 Object-Based Maps

A second direction concerns the task of segmentation (in conjunction with learning): identifying which objects the sensor data came from, and (simultaneously) learning the object properties. We extended existing robot mapping techniques by incorporating the notions of objects and classes into the learning framework. Such modeling allows us to naturally prefer models in which objects of the same class have similar properties, and leverages generalization and prediction.

As a first step [1], we implemented a system learning 2D maps of office environments with two kinds of objects (walls and doors) from real laser scanner and camera robot data. The original model encodes general knowledge such as “doors open and close over time while walls remain static” and “doors usually share width and orientation”, and is optimized by a novel instance of the EM algorithm which yields a map in terms of 2D segments corresponding to doors and walls. Our experiments demonstrate the use of object property generalization and prediction. Even though we originally don’t know how walls and doors look in a particular environment, our method finds a few doors based on their motion, then generalizes from the newly acquired color information to find doors that never moved at all.

2.3 3D Shape Models

Our final direction addresses the problem of constructing 3D models for the shapes of complex objects, specifically objects whose shape is not static. We view the ability to model shape variation both as necessary in itself, in order to deal with objects such as people or chairs, and as a key building block in constructing class-based models of 3D objects (where shape variability is almost always the case).

In the context of this project, we have made significant progress on two tasks. In the first [3], we developed an unsupervised algorithm for registering 3D surface scans of an object undergoing significant deformations. Our algorithm does not use markers, nor does it assume prior knowledge about object shape, the dynamics of its deformation, or scan alignment. The algorithm registers two meshes by optimizing a joint probabilistic model over all point-to-point correspondences between them. This model enforces preservation of local mesh geometry, as well as more global constraints that capture the preservation of geodesic distance between corresponding point pairs. The algorithm applies even when one of the meshes is an incomplete range scan; thus, it can be used to automatically fill in the remaining surfaces for this partial scan, even if those surfaces were previously only seen in a different configuration. We evaluated the algorithm on several real-world

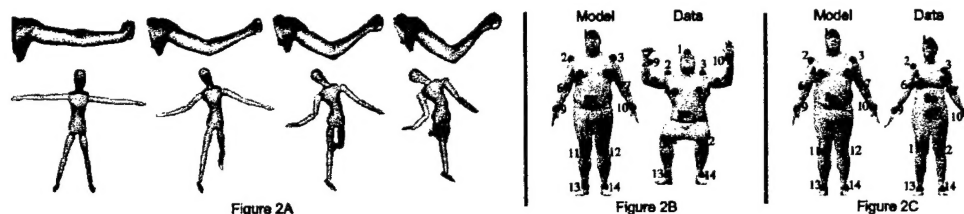


Figure 7: Sample results from our unsupervised correspondence algorithm. (2A) Automatic interpolation between two scans of an arm and a wooden puppet. (2B) Registration results on two scans of the same man sitting and standing up (select points were displayed) (2C) Registration results on scans of a larger man and a smaller woman. The algorithm is robust to small changes in object scale.

datasets, and demonstrated good results in the presence of significant movement of articulated parts and non-rigid surface deformation.

In our second task [2], we considered the task of automatically determining the decomposition of an articulate object into its constituent approximately rigid parts. The input of our algorithm is a set of meshes corresponding to different configurations of an articulated object, registered using the algorithm described above. Our algorithm automatically recovers the part decomposition, the location of the parts in the different object instances, and the articulated object skeleton linking the parts. Our algorithm segments the mesh surfaces using a graphical model that captures the spatial contiguity of parts. The segmentation is done using the EM algorithm, iterating between finding a decomposition of the object into rigid parts, and finding the location of the parts in the object instances. Although the graphical model is densely connected, the object decomposition step can be performed optimally and efficiently, allowing us to identify a large number of object parts while avoiding local maxima.

References

- [1] D. Anguelov, D. Koller, E. Parker, and S. Thrun. Detecting and modeling doors with mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.
- [2] D. Anguelov, H.-C. Pang, D. Koller, P. Srinivasan, and S. Thrun. Recovering articulated object models from 3D range data. In *Proceedings of the Conference on Uncertainty in AI (UAI)*, 2004.
- [3] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, H.-C. Pang, and J. Davis. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *Advances in Neural Information Processing Systems (NIPS)*, 2004. To appear.
- [4] X. Boyen and D. Koller. Approximate learning of dynamic models. In *Advances in Neural Information Processing Systems (NIPS)*, 1998.
- [5] X. Boyen and D. Koller. Exploiting the architecture of dynamic systems. In *Proceedings of the 16th National Conference on Artificial Intelligence*, 1999.

- [6] Xavier Boyen, Nir Friedman, and Daphne Koller. Discovering the hidden structure of complex dynamic systems. In *Proceedings of the 15th Annual Conference on Uncertainty in AI (UAI)*, pages 91–100, 1999.
- [7] Xavier Boyen and Daphne Koller. Tractable inference for complex stochastic processes. In *Proceedings of the 14th Annual Conference on Uncertainty in AI (UAI)*, pages 33–42, 1998.
- [8] U. Lerner, B. Moses, M. Scott, S. McIlraith, and D. Koller. Monitoring a complex physical system using a hybrid dynamic Bayes net. In *Proceedings of the 18th Annual Conference on Uncertainty in AI (UAI)*, pages 301–310, 2002.
- [9] U. Lerner and R. Parr. Inference in hybrid networks: Theoretical limits and practical algorithms. In *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, 2001.
- [10] U. Lerner, R. Parr, D. Koller, and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proceedings of the 17th National Conference on Artificial Intelligence*, pages 531–537, 2000.
- [11] U. Lerner, E. Segal, and D. Koller. Exact inference in networks with discrete children of continuous parents. In *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, pages 319–238, 2001.
- [12] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the 18th National Conference on Artificial Intelligence*, 2002.
- [13] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.
- [14] S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research*, 2004. To Appear.